

ESTUDIO COMPARATIVO DE LOS COEFICIENTES DE CORRELACION OBTENIDOS EN TABLAS DE CONTINGENCIA DE 2X2 (II): MUESTRAS DE DIFERENTE TAMAÑO

por O. LEON y F. J. TEJEDOR
Universidad Autónoma y Complutense de Madrid

El presente artículo es la segunda parte del publicado con el mismo título general en el número anterior de esta misma revista, referido entonces a muestras del mismo tamaño.

Tanto la terminología como los conceptos básicos implicados están allí definidos por lo que rogamos al lector tenga la amabilidad de consultarlo, si es que no lo ha hecho ya.

Las preguntas que dan lugar a esta segunda parte del trabajo podrían formularse en los siguientes términos: ¿se mantendrá la tendencia expuesta (y por tanto las conclusiones) de la primera parte del trabajo respecto a los distintos coeficientes de correlación si las muestras tuvieran diferentes tamaños? ¿Qué repercusiones puede tener en la investigación en ciencias humanas el hecho de trabajar con muestras de diferente tamaño?

Como entonces, trataremos de justificar la respuesta basándonos en situaciones simuladas que nos ayuden a determinar tendencias, similitudes, etc.

2. Estudio de la variación de los coeficientes de correlación en tablas 2 × 2 con muestras de tamaños diferentes.

Seguimos considerando el mismo ejemplo en lo que se refiere a variables, categorías y disposición de frecuencias en la tabla:

	aceptación	rechazo
hombres	A	B
mujeres	C	D

El tamaño total de las muestras, $N = 80$, se tomará como base para irlo modificando adecuadamente.

Seguiremos, cuando sea preciso, recurriendo al parámetro OTE (diferencia total entre frecuencias empíricas por muestras).

2.1. Incremento del tamaño de una muestra con diferencia nula entre las frecuencias; la otra muestra se mantiene en una relación constante

Mantendremos la muestra de hombres en una relación constante determinada, por ejemplo

30	10
----	----

La muestra de mujeres, con diferencia nula entre las frecuencias empíricas de ambas casillas, se irá incrementando paulatinamente, en razón a una constante de transformación k , a partir del tamaño de las muestras tomado como base en la primera parte del trabajo.

Así pues, la tabla de frecuencias de partida será:

caso 1

30	10	40	$k = 1$	$\phi = 0,2582$
20	20	40	$\chi^2 = 5,3333$	$O_c = 0,25$
50	30	80	$C_s = 0,3535$	$W = 0,2679$
			$Q = 0,5$	
			$r_c = 0,4086$	

caso 2

30	10	40	$k = 1,5$	$\phi = 0,25$
30	30	60	$\chi^2 = 6,25$	$O_c = 0,24$
60	40	100	$C_s = 0,343$	$W = 0,2679$
			$Q = 0,5$	
			$r_c = 0,4086$	

caso 3

30	10	40	$k = 2$	$\phi = 0,2392$
40	40	80	$\chi^2 = 6,8638$	$O_c = 0,2223$
70	50	120	$C_s = 0,3289$	$W = 0,2679$
			$Q = 0,5$	
			$r_c = 0,4086$	

caso 8

30	10	40	$k = 4,5$	$\varphi = 0,1937$
			$\chi^2 = 8,2566$	$Q = 0,1487$
90	90	180	$C = 0,2690$	$W = 0,2679$
			$Q = 0,5$	
120	100	220	$r = 0,4086$	

En la tabla 5 se recogen los datos del conjunto de casos que utilizaremos para la obtención de la gráfica correspondiente.

Tabla 5

caso	k							
1	1	5,3333	0,2582	0,3535	0,25	0,5	0,2679	0,4086
2	1,5	6,25	0,25	0,3430	0,24	0,5	0,2679	0,4086
3	2	6,8638	0,2392	0,3289	0,2223	0,5	0,2679	0,4086
4	2,5	7,2860	0,2281	0,3145	0,2040	0,5	0,2679	0,4086
5	3	7,6190	0,2182	0,3015	0,1875	0,5	0,2679	0,4086
6	3,5	7,8794	0,2092	0,2896	0,1722	0,5	0,2679	0,4086
7	4	8,0808	0,2010	0,2787	0,16	0,5	0,2679	0,4086
8	4,5	8,2566	0,1937	0,2690	0,1487	0,5	0,2679	0,4086
•	•							
•	•							
•	•							

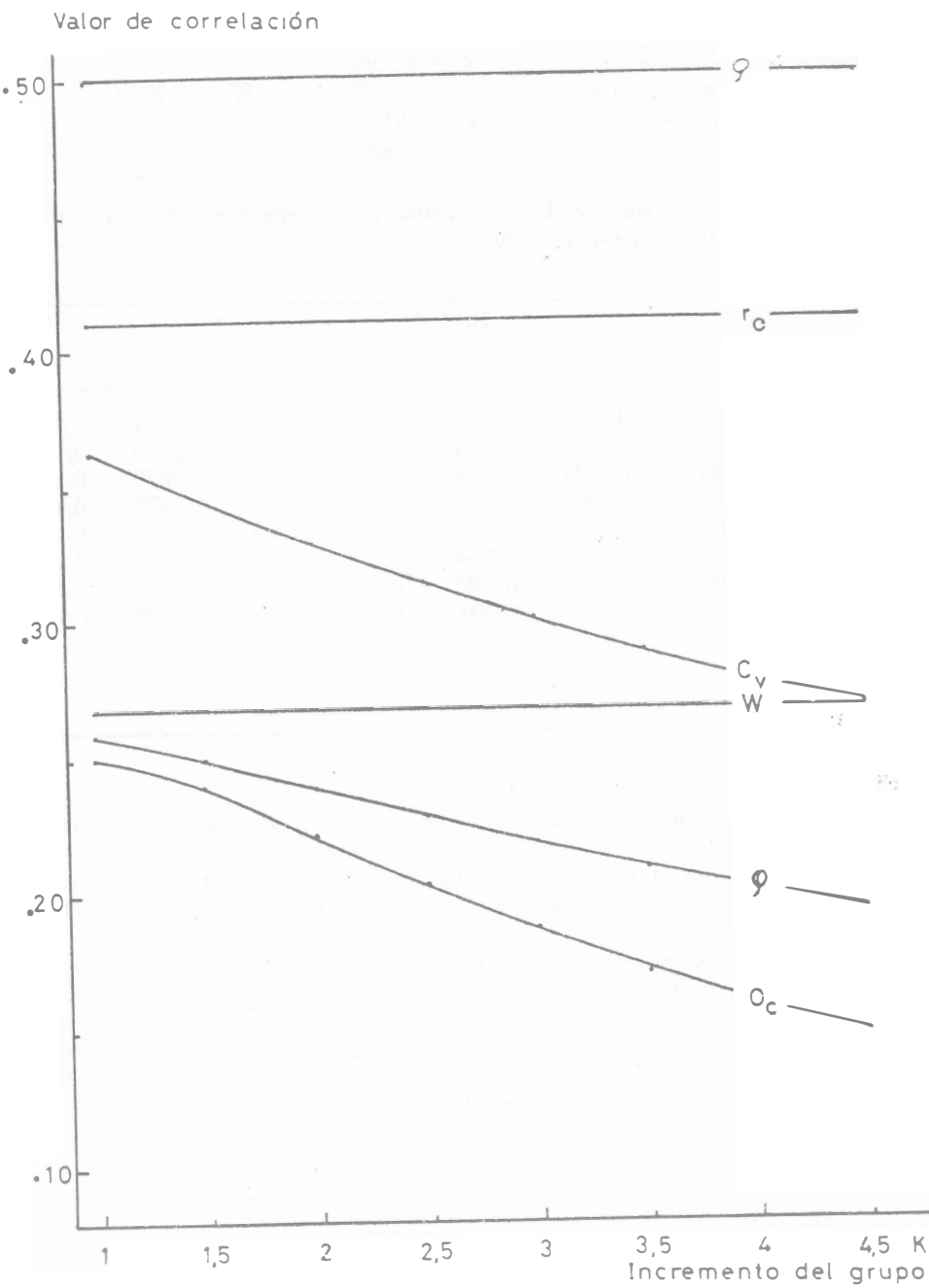
Como la variación se ha producido por los sucesivos valores que ha tomado k, haremos que ésta figure en el eje de abscisas en tanto que los valores de correlación los situaremos en la ordenada. En esta representación (gráfica número 5) sólo contemplamos el intervalo 0,10-0,50 para los valores de correlación ya que los coeficientes calculados varían dentro de esos límites.

Los resultados pueden parecernos de alguna manera, y a primera vista, sorprendentes; y ello porque son los coeficientes Q, W y r los que permanecen constantes cuando aumentamos el tamaño de una de las muestras mientras la otra se mantiene en una relación determinada, lo que puede hacernos pensar que serían éstos los más rigurosos.

Preocupados por esa primera impresión de la gráfica 5, observamos sin dificultades que esa igualdad de Q, W y r es simplemente consecuencia de su fórmula. Por ejemplo, para Q, vemos que:

$$Q = \frac{AD - BC}{AD + BC}$$

gráfica 5



$$Q_1 = \frac{AkD - BkC}{AkD + BkC} = \frac{k (AD - BC)}{k (AD + BC)} = \frac{AD - BC}{AD + BC} = Q$$

Para W y r se harían demostraciones similares.

Por otra parte, nos fijamos que la variación de ϕ , C_r y Q_r tiene una explicación relativamente satisfactoria, en el sentido de que estamos comparando dos muestras y al variar el tamaño de una se modifica la relación de proporcionalidad entre las frecuencias, que en definitiva es lo que tratan de reflejar los coeficientes de correlación. Es decir, fijémonos en los casos 2 y 5, por ejemplo.

En el caso 2, el número de mujeres que está de acuerdo con la participación de los padres en la dirección del Centro es el mismo que el de hombres (razón 1 : 1) mientras que el número de mujeres que no está de acuerdo está en la proporción 3 : 1.

En el caso 5, las proporciones entre las frecuencias son 2 : 1 y 6 : 1, respectivamente.

Entas diferencias entre las frecuencias explican, a nuestro entender, la variación que muestran los coeficientes ϕ , C_r y Q_r , aunque no podemos por menos que dudar respecto a la conveniencia de que las muestras tengan diferente tamaño.

2.2. Incremento del tamaño de una muestra con una diferencia dada; la otra muestra se mantiene en una relación constante

Supone una generalización del caso anterior. Utilizamos los mismos valores de k y tomaremos como tabla base de frecuencias la siguiente:

30	10
10	30

Transformaremos (en base a k) las frecuencias de la muestra de mujeres lo que nos permitirá ir creando por simulación los diferentes casos.

caso 1

30	10	40	$k = 1$	$\phi = 0,5$
10	30	40	$\chi^2 = 20$	$Q_r = 0,5$
40	40	80	$C_r = 0,634$	$W = 0,50$
			$Q = 0,80$	
			$r_c = 0,7071$	

caso 2

30	10	40	$k = 1,5$	$\phi = 0,4924$
15	45	60	$\chi^2 = 24,2424$	$Q_r = 0,48$
45	55	100	$C_r = 0,6047$	$W = 0,50$
			$Q = 0,80$	
			$r_c = 0,7071$	

caso 3

30	10	40	$k = 2$	$\phi = 0,4780$
			$\chi^2 = 27,4141$	$\alpha = 0,4443$
20	60	80	$C_v = 0,6099$	$W = 0,50$
			$Q = 0,80$	
50	70	120	$r_c = 0,7071$	

caso 8

30	10	40	$k = 4,5$	$\phi = 0,4067$
			$\chi^2 = 36,3940$	$\alpha = 0,2974$
45	135	180	$C_v = 0,5328$	$W = 0,50$
			$Q = 0,80$	
75	145	220	$r_c = 0,7071$	

En la tabla 6 se recogen los datos del conjunto de casos que utilizaremos para la obtención de la gráfica.

Tabla 6

caso	k							
1	1	20	0,50	0,6340	0,50	0,80	0,50	0,7071
2	1,5	24,2424	0,4924	0,6247	0,48	0,80	0,50	0,7071
3	2	27,4141	0,4780	0,6099	0,4443	0,80	0,50	0,7071
4	2,5	29,9662	0,4626	0,5938	0,4083	0,80	0,50	0,7071
5	3	32	0,4472	0,5773	0,375	0,80	0,50	0,7071
6	3,5	33,7349	0,4329	0,5618	0,3458	0,80	0,50	0,7071
7	4	35,1648	0,4193	0,5469	0,32	0,80	0,50	0,7071
8	4,5	36,3940	0,4067	0,5328	0,2974	0,80	0,50	0,7071
.	.							
.	.							
.	.							

La gráfica 6 es la asociada a estos datos.

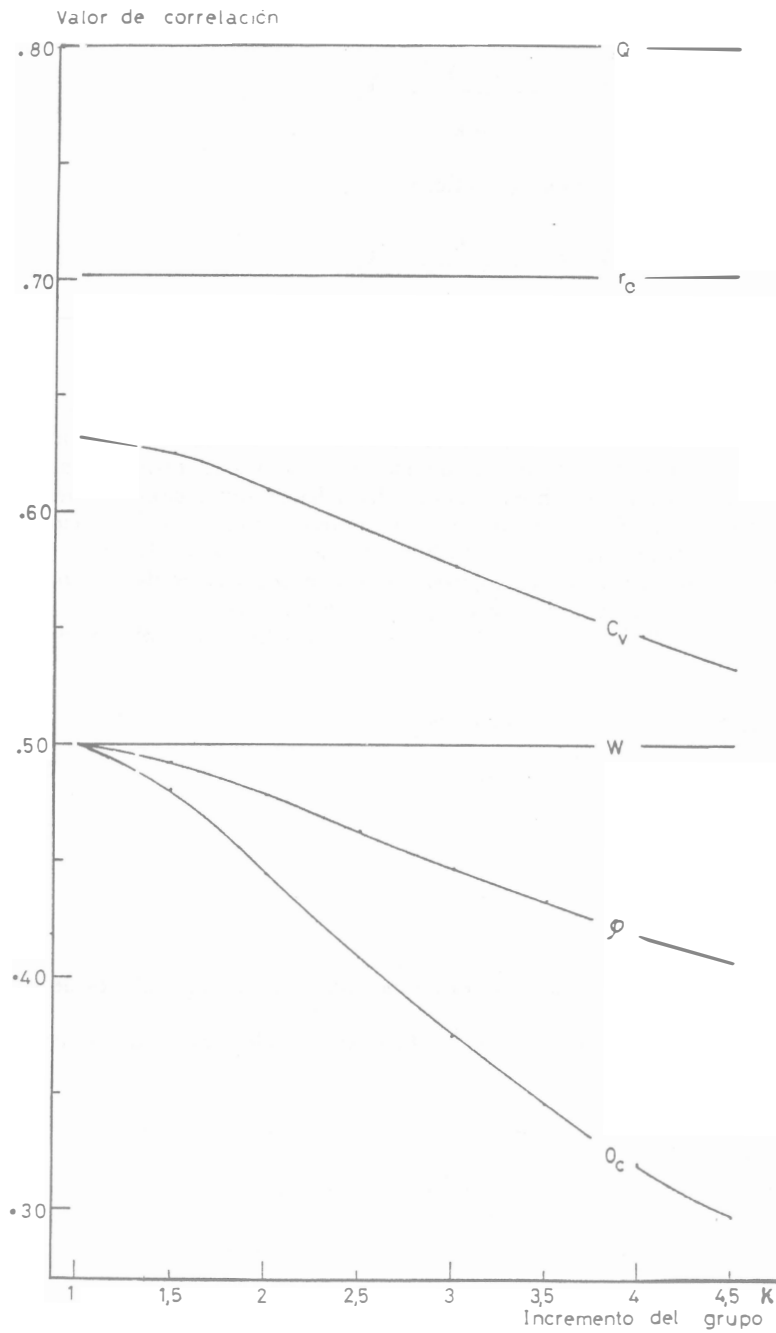
Podemos ver que al igual que en la gráfica 5 se siguen manteniendo constantes los valores de Q , W y r_c , mientras que ϕ , C_v y α disminuyen, como ocurría entonces, a medida que k aumenta.

Quizá convenga destacar que la caída de los valores ϕ , C_v y α , sobre todo de este último, es más rápido que en la gráfica 5; ello nos hace pensar que al ser mayor la diferencia entre las frecuencias del grupo aumentado, la correlación disminuye proporcionalmente al ir aumentando el tamaño de la muestra (todo ello, claro está, para tamaño y diferencia fijos en la otra muestra).

Parece claro que los comentarios referidos a la gráfica 5 son aplicables a la gráfica 6, ya que, como dijimos, se trata de una generalización de las observaciones entonces expuestas.

La formalización de estas generalización se muestra en el trabajo original en las

gráfica 6



páginas 461-504. Así, por ejemplo, la razón de variación de α al ampliar una de las muestras viene dada por:

$$\alpha_1 = \frac{(1+k)^2}{4k} \alpha_2$$

O si se prefiere,

$$\alpha_2 = \frac{4k}{(1+k)^2} \alpha_1$$

2.3. Variación aleatoria

Como último estudio nos hubiera gustado presentar las variaciones ocurridas en los diferentes coeficientes cuando las frecuencias de las tablas estan fuera de "las condiciones de laboratorio" que nosotros hemos ido simulando. Esto sólo será posible cuando, utilizando el ordenador, podamos establecer pautas adecuadas de simulación. Es nuestro objetivo para un trabajo posterior. Por ahora, terminaremos el estudio de las variaciones presentando una nueva situación, con $N = 80$, manteniendo el valor $DTE = 10$ y desarrollando ordenadamente todos los casos, pasando de la máxima acumulación en una de las casillas (y por tanto mínima en la otra) a la situación inversa.

Al haber fijado el valor $DTE = 10$, las diferencias entre empíricas en cada una de las muestras podrá ser:

1. ^a muestra	2. ^a muestra
10	0
9	1
8	2
7	3
6	4
5	5

(los pares 4-6, 3-7..., etc., no se consideran porque no son más que repeticiones de sus simétricos).

Así pues, para las diferencias 10-0, la serie de tablas, recordando que en todos los casos $DTE = 10$, sería:

DE ₁ = 10	<table><tr><td>45</td><td>35</td></tr><tr><td>0</td><td>0</td></tr></table>	45	35	0	0	<table><tr><td>44</td><td>34</td></tr><tr><td>1</td><td>1</td></tr></table>	44	34	1	1	<table><tr><td>43</td><td>33</td></tr><tr><td>2</td><td>2</td></tr></table>	43	33	2	2	...	<table><tr><td>10</td><td>0</td></tr><tr><td>35</td><td>35</td></tr></table>	10	0	35	35
45	35																				
0	0																				
44	34																				
1	1																				
43	33																				
2	2																				
10	0																				
35	35																				
DE ₂ = 0																					

De forma similar se desarrollarían las series de tablas para los pares 9-1, 8-2,...

Presentamos los valores que los diversos coeficientes toman en algunas de las tablas de las distintas series, aquellas que hemos utilizado para la obtención de las gráficas.

a) Par de diferencias 10-0:

caso 1

45	35	80
0	0	0
45	35	80

DTE = 10

$$\chi^2 = \phi = C_v = \alpha = Q = W = r_c = 0$$

caso 2

42	32	74
3	3	6
45	35	80

DTE = 10

$\chi^2 =$

$C_v = 0,0507$

$Q = 0,1190$

$r_c = 0,1064$

$\phi = 0,0359$

$\alpha = 0,0187$

$W = 0,0679$

caso 3

38	28	66
7	7	14
45	35	80

DTE = 10

$\chi^2 = 0,2584$

$C_v = 0,0802$

$Q = 0,1316$

$r_c = 0,1194$

$\phi = 0,0568$

$\alpha = 0,0437$

$W = 0,0762$

caso 10

10	0	10
35	35	70
45	35	80

DTE = 10

$\chi^2 = 8,8889$

$C_v = 0,4472$

$Q = W = r_c = 1$

$\phi = 0,3333$

$\alpha = 0,2187$

b) Par de diferencias 9-1:

caso 1

44	35	79
0	1	1
44	36	80

DTE = 10

$\chi^2 = 0,8285$

$C_v = 0,1432$

$Q = W = r_c = 1$

$\phi = 0,1018$

$\alpha = 0,0275$

caso 10

9	0	9
35	36	71
44	36	80

DTE = 10
 $\chi^2 = 8,2970$
 $C_v = 0,4335$
 $Q = W = r_c = 1$

$\varphi = 0,3220$
 $\alpha = 0,2025$

c) Par de diferencias 8-2:

caso 1

43	35	78
0	2	2
43	37	80

DTE = 10
 $\chi^2 = 2,3817$
 $C_v = 0,2405$
 $Q = W = r_c = 1$

$\varphi = 0,1725$
 $\alpha = 0,0537$

caso 10

8	0	8
35	37	72
43	37	80

DTE = 10
 $\chi^2 = 7,6485$
 $C_v = 0,4178$
 $Q = W = r_c = 1$

$\varphi = 0,3092$
 $\alpha = 0,1850$

d) Par de diferencias 7-3:

caso 1

42	35	77
0	3	3
42	38	80

DTE = 10
 $\chi^2 = 3,4450$
 $C_v = 0,2873$
 $Q = W = r_c = 1$

$\varphi = 0,02075$
 $\alpha = 0,0787$

caso 10

7	0	7
35	38	73
42	38	80

DTE = 10
 $\chi^2 = 7,2153$
 $C_v = 0,4068$
 $Q = W = r_c = 1$

$\varphi = 0,3003$
 $\alpha = 0,1662$

e) Par de diferencias 6-4:

caso 1

41	35	76
0	4	4
41	39	80

$$\begin{aligned} \text{DTE} &= 10 \\ \chi^2 &= 4,4264 \\ C_v &= 0,3238 \\ Q &= W = r_c = 1 \end{aligned}$$

$$\begin{aligned} \phi &= 0,2352 \\ \alpha &= 0,1025 \end{aligned}$$

caso 10

6	0	6
35	39	74
41	39	80

$$\begin{aligned} \text{DTE} &= 10 \\ \chi^2 &= 6,1701 \\ C_v &= 0,3784 \\ Q &= W = r_c = 1 \end{aligned}$$

$$\begin{aligned} \phi &= 0,2777 \\ \alpha &= 0,1462 \end{aligned}$$

f) Par de diferencias 5-5:

caso 1

40	35	75
0	5	5
40	40	80

$$\begin{aligned} \text{DTE} &= 10 \\ \chi^2 &= 5,3333 \\ C_v &= 0,3536 \\ Q &= W = r_c = 1 \end{aligned}$$

$$\begin{aligned} \phi &= 0,2582 \\ \alpha &= 0,1250 \end{aligned}$$

caso 10

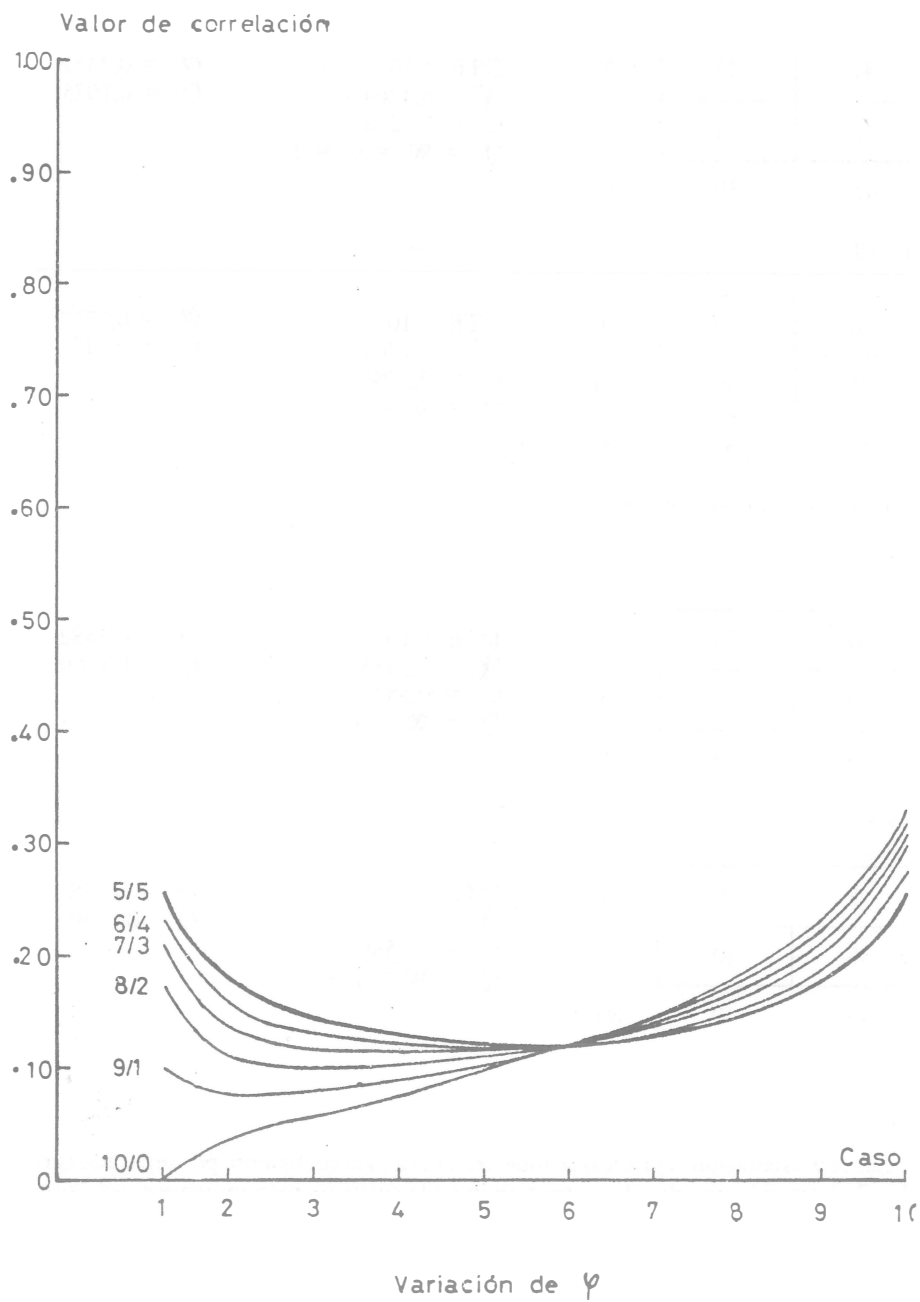
5	0	5
35	40	75
40	40	80

$$\begin{aligned} \text{DTE} &= 10 \\ \chi^2 &= 5,3333 \\ C_v &= 0,3536 \\ Q &= W = r_c = 1 \end{aligned}$$

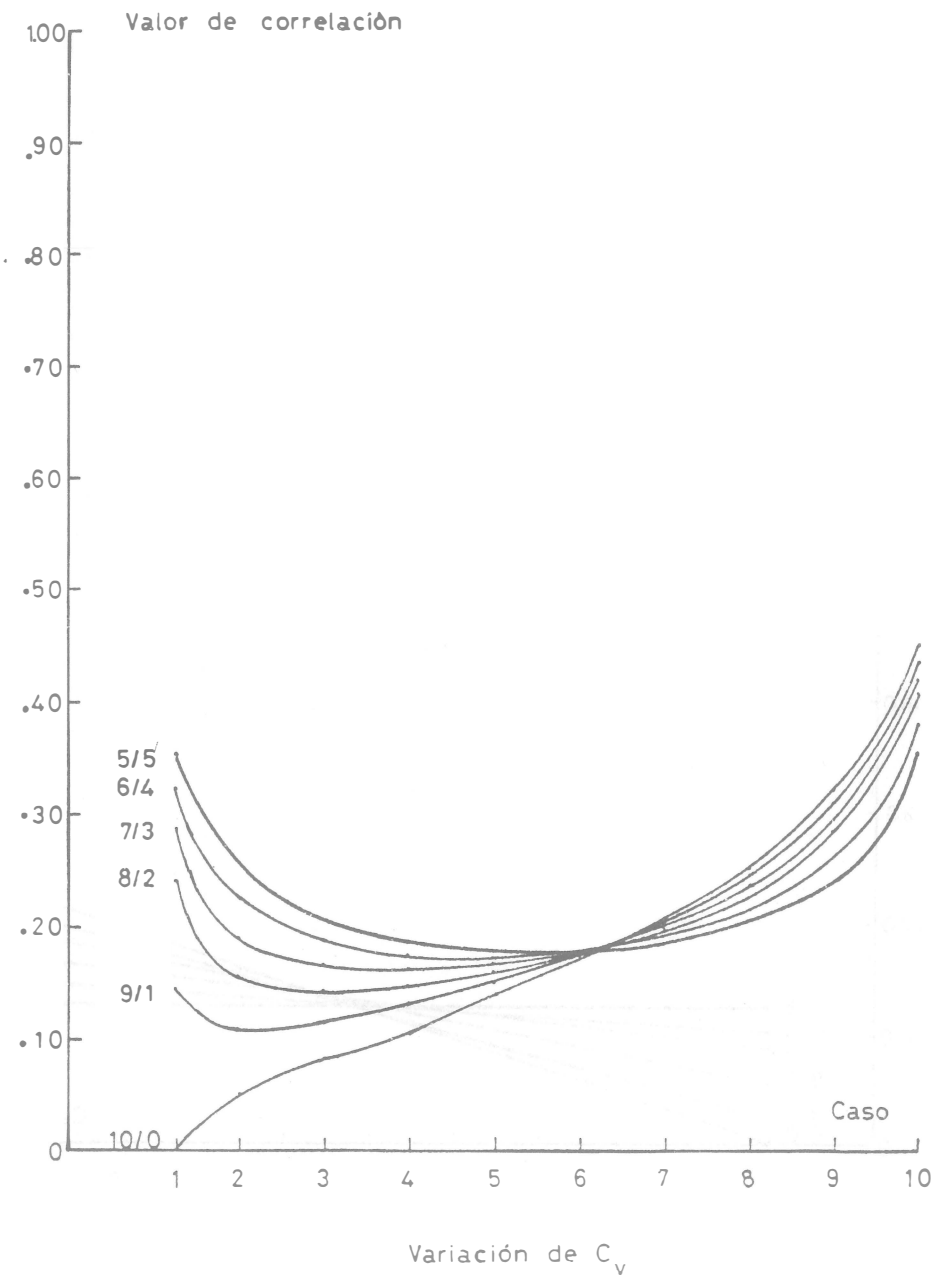
$$\begin{aligned} \phi &= 0,2582 \\ \alpha &= 0,1250 \end{aligned}$$

Las representaciones gráficas se muestran para cada coeficiente por separado (gráficas 7.1 a 7.6) en cada una de ellas representamos la curva correspondiente a cada situación estudiada (los distintos pares de diferencias); sobre la abscisa llevamos los diferentes casos presentados (10) y sobre la ordenada los valores que toma la correlación; no hemos acotado los valores de la ordenada para poder tener referencias comparativas fidedignas.

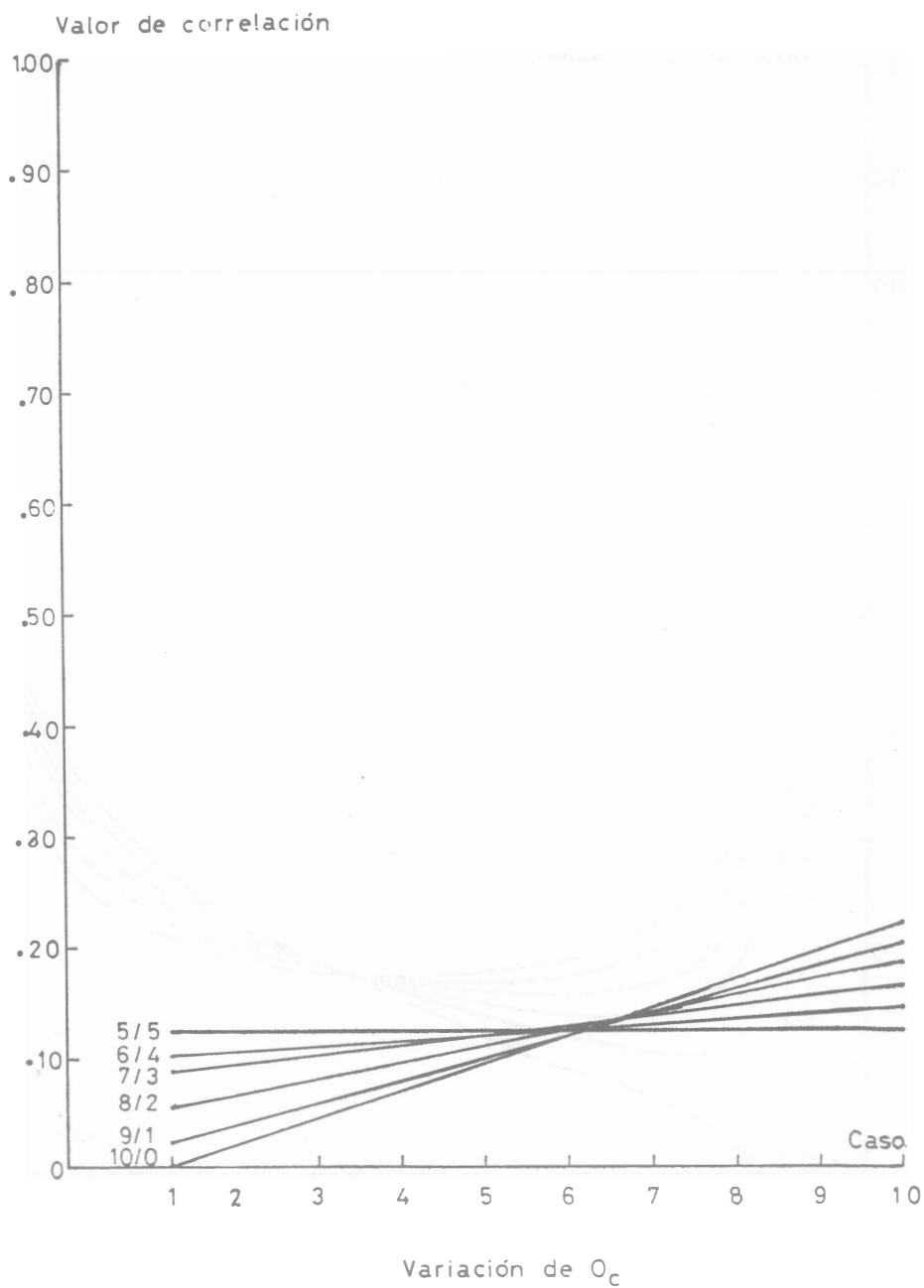
gráfica 7.1



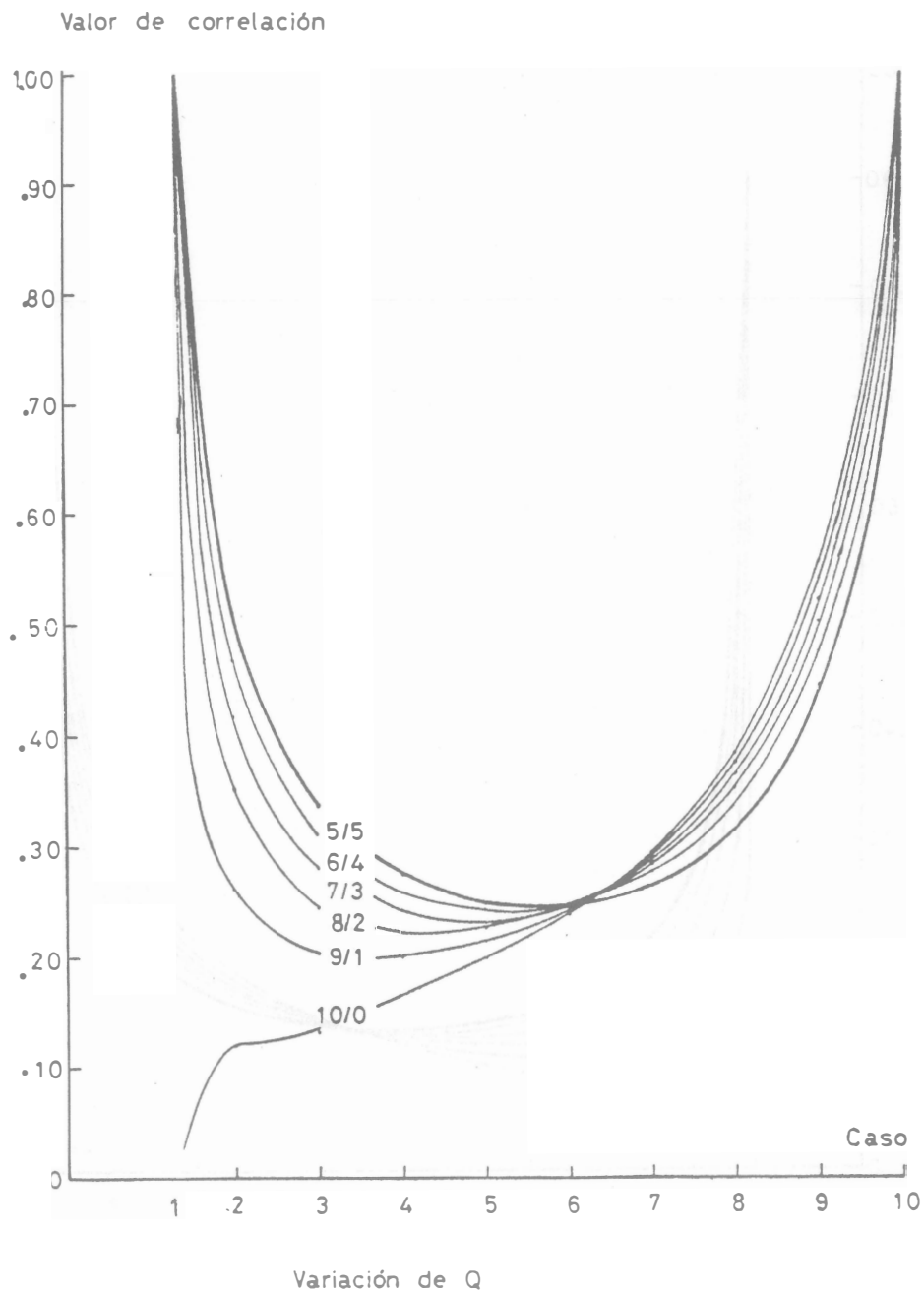
gráfica 7.2



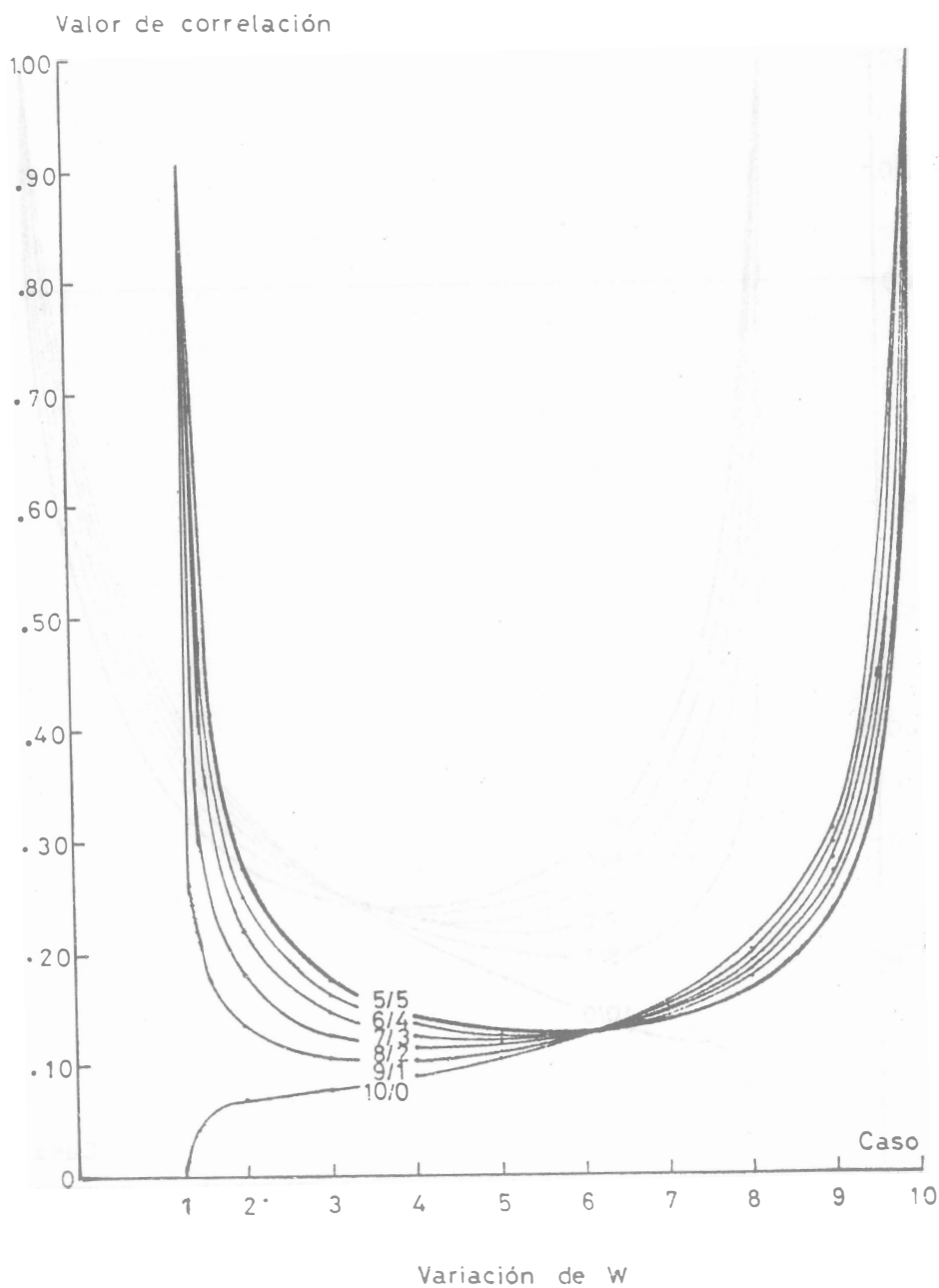
gráfica 7.3



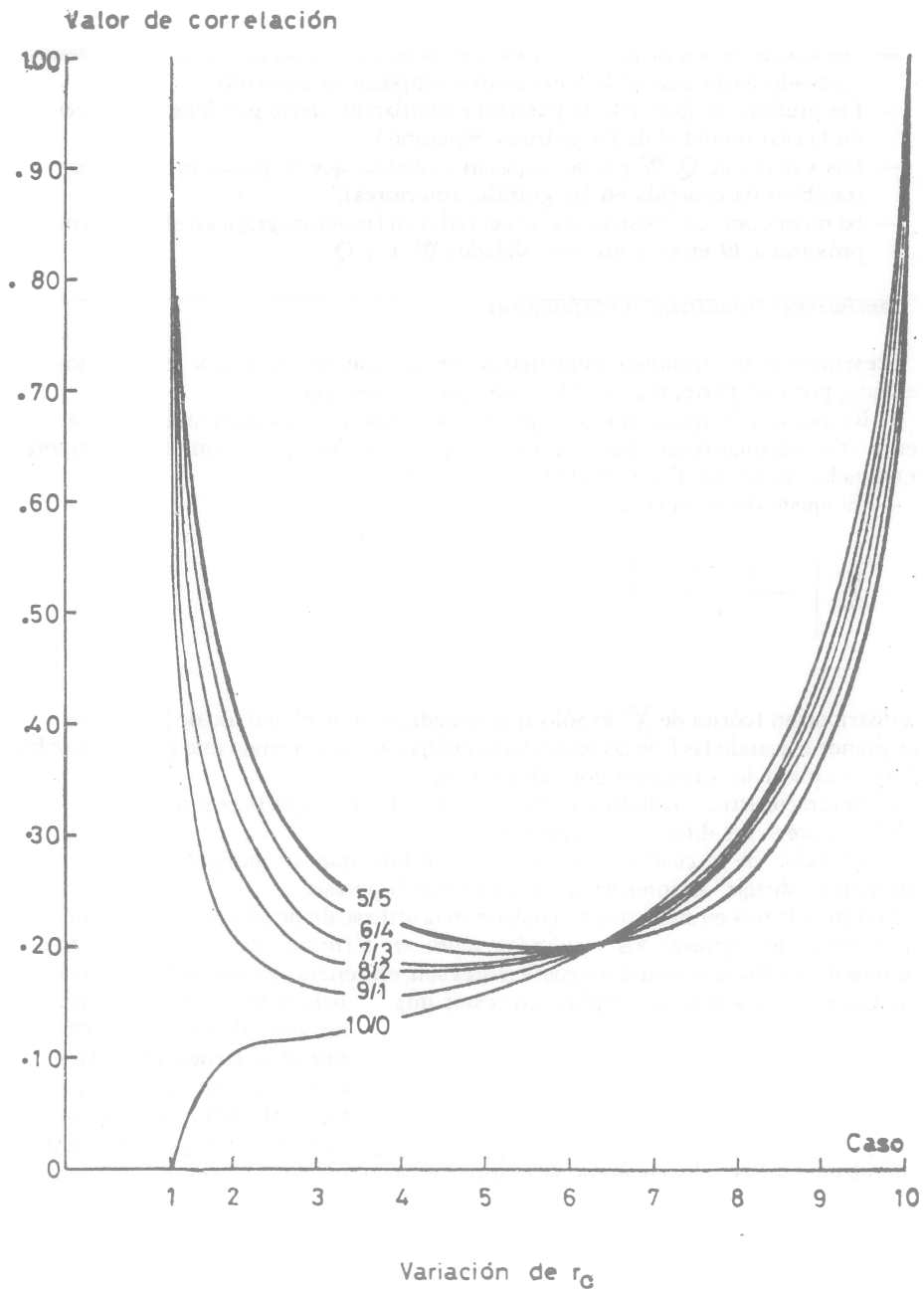
gráfica 7.4



gráfica 7.5



gráfica 7.6



De la observación de las gráficas 7.1 a 7.6 el lector puede sacar, creemos, sus propias conclusiones. Nosotros destacamos únicamente algunas de las que consideramos más notorias:

- Las curvas, no monótonas, se cruzan en torno al caso 6 (caso intermedio).
- Los trazos de los diversos coeficientes presentan (excepto para O_c) un descenso atenuado hasta que en la zona central empieza su aumento.
- Las gráficas de φ y de C se parecen y guardan un cierto paralelismo (ha ocurrido en la casi totalidad de las gráficas expuestas).
- Los valores de Q , W y r_c se disparan a medida que se producen casos extremos (también ha ocurrido en las gráficas anteriores).
- Se mantienen las "posiciones" observadas en las demás gráficas: el coeficiente más próximo a φ es O_c y los más alejados W , r_c y Q .

3. Conclusiones estadístico-metodológicas

Presentamos un resumen esquemático de las conclusiones. Las referiremos a ji-cuadrado, por una parte, y a los diferentes coeficientes, por otra.

a) Respecto a la aplicación de la prueba de ji-cuadrado entendemos exige tener en cuenta tres peculiaridades para realizar una mejor interpretación de las hipótesis contrastadas mediante dicho estadístico:

- El ajuste de la función

$$\sum \left[\frac{(f_e - f_t)^2}{f_t} \right]$$

a la distribución teórica de χ^2 es sólo aproximada; cuando el tamaño de la muestra no es muy grande o cuando las f_t de las respectivas casillas se van haciendo más pequeñas se hace necesario aplicar la corrección por continuidad,

- la característica cuadrática del término $(f_e - f_t)^2$ hace que se potencien más —quizá excesivamente— las diferencias mayores,

- el valor de ji-cuadrado aumenta a medida que se aumenta N ; ello puede permitirnos "dirigir" la interpretación de ciertas hipótesis.

Recomendamos en suma mucho cuidado en la utilización de esta prueba; creemos en la conveniencia de expresar los resultados tanto en términos de ji-cuadrado como en términos de correlación, ya que los coeficientes son, en general, menos sensibles al tamaño de la muestra; esta postura implica un restar importancia al hecho de la significación estadística (aportación de χ^2) y potenciar el cuestionamiento de la relación entre las categorías de las variables en función del tamaño de las muestras; en definitiva, potenciar la significación psicológica, pedagógica, sociológica del estudio en detrimento de la significación únicamente estadística como juicio supremo y definitivo, y ello porque, una vez más hemos de decirlo, pensamos que la consecución de esta significación puede manipularse en base al tamaño de la muestra elegida.

b) Respecto a los diferentes coeficientes, aunque ya hemos presentado a lo largo del trabajo algunas de las conclusiones, especialmente en el punto 1.5, a título de resumen nos parece conveniente destacar:

- En el caso de una variación uniforme, φ , W y O_c responden a una ecuación lineal

mientras que C_v , Q y r_c describen ramas parabólicas más infladas por encima de su punto medio.

— Cuando la distribución de las frecuencias no es uniforme parece que es O_c el coeficiente que mejor refleja la correlación que intuitivamente cabría esperar de la disposición de los datos; ϕ se muestra más próximo a O_c en tanto que Q , W y r_c se disparan hacia 1 para diferencias extremas entre las frecuencias.

— Nos parece más adecuado ϕ que C_v para tablas de 2×2 ; C_v siempre da valores más altos que ϕ y sus gráficas respectivas mantienen un cierto paralelismo.

— Cuando hay alguna frecuencia de cero no procede la aplicación de Q , W y r_c ; tampoco es recomendable su utilización cuando una de las muestras presenta máxima dependencia o cuando una muestra presenta frecuencias iguales en cada una de las categorías de la variable en estudio.

— El hecho de aumentar el tamaño de una muestra, manteniendo la relación constante, hace que χ^2 aumente y que ϕ , C_v y O_c disminuyan, siendo esta disminución inversamente proporcional al aumento del tamaño de la muestra y directamente proporcional a la dependencia entre las muestras.

— Para una relación dada entre dos muestras respecto a una variable dicotomizada siempre obtendremos valores de correlación ϕ menores cuando los tamaños de las muestras sean diferentes frente al caso de tamaños iguales.

Lo mismo podemos decir respecto a C_v y respecto a O_c

Esta observación la consideramos tan importante que nos atrevemos a sospechar que cuando medimos el grado de independencia/correlación mediante el coeficiente ϕ y lo hacemos sobre el tamaño de muestras diferentes, las conclusiones a las que llegamos pueden ser falsas; y ello al margen de la probabilidad de error que conlleva el contraste de cualquier hipótesis estadística.

Es frecuente en la investigación en ciencias humanas que se comparen los grupos experimental y de control; las propias dificultades de la experimentación suelen traducirse en un desequilibrio notorio entre los tamaños de los respectivos grupos (mucho mayor el grupo control); en estas condiciones, la aplicación de χ^2 tiende a mostrar diferencias inexistentes, debidas más a la estructura de su propia función que a la influencia de la variable que intentamos estudiar; lo mismo ocurre respecto a los diferentes coeficientes de correlación, aunque como ya hemos mencionado, en menor medida.

En definitiva, aconsejamos que para los contrastes entre muestras a las que aplicamos χ^2 y para el estudio de la relación en tablas de contingencia de 2×2 , los tamaños de las muestras sean iguales, en la seguridad de que a medida de que sean más distintos la interpretación de hipótesis de significación de diferencias y la expresión de relación entre las variables es menos consistente, menos fiable.

Dirección del autor: F. J. Tejedor, O. G. León, c/ Oña, 9, Madrid-34

BIBLIOGRAFIA

- ALTHAM, P. M. E. (1970): The measurement of association of rows and columns for $r \times s$ contingency tables, *Journal Research Statistic Social*, Serie B 32, pp. 63-73.
- BHAPKAR, V. P. y SOMES, G. W. (1977): Distribution of Q when testing equality of matched proportions, *Journal American Statistic Association*: 72. pp. 658-661.
- BRADLEY, J. V. (1968): *Distribution-free statistical test* (Englewood Cliffs, N. J., Prentice Hall).

- (1972) Nonparametrics statistics, en WIRK, R. E. (Ed). *Statistical Issues* (Monterrey. California, Brooks Cole), pp. 329-338.
- BROWN, M. B. y BENEDETTI, J. K. (1977): Sampling behaviour of test for correlation in two-way contingency tables, *Journal American Statistic Association*: 72, pp. 309-315.
- CAMILLI, G. y HOPKINS, K. D. (1978): Applicability of chi-square to 2×2 contingency tables with small expected cell frequencies, *Psychological Bulletin*: 85, pp. 163-167.
- CHIACCHIERINI, R. P. y ARNOLD, J. C. (1977): A two sample test for independence in 2×2 contingency tables with both margins subject to misclassification, *Journal American Statistic Association*: 72, pp. 170-174.
- CONOVER, W. J. (1971): *Practical nonparametrics statistics* (New York, Wiley).
- EDWARDS, A. L. (1950): On the use and misuse of the chi-square test, *Psychological Bulletin*: 47, pp. 341-346.
- FLEISS, J. L. y EVERITT, B. S. (1971): Comparing the marginal totals of square contingency tables, *British Journal Mathematical Statistic Psychological*: 24, pp. 117-123.
- GAIL, M. y MANTEL, N. (1977): Counting the number of $r \times s$ contingency tables with fixed margins. *Journal American Statistic Association*: 72, pp. 859-862.
- IVES, K. H. y GIBBONS, J. D. (1967): A correlation measure for nominal data, *American Statistic*: 21, pp. 16-17.
- LEWIS, D. y BURKE, C. J. (1949): The use and misuse of the chisquare test, *Psychological Bulletin*: 46, pp. 433-489.
- (1950): Further discussion of the use and misuse of the chi-square test, *Psychological Bulletin*: 47, pp. 433-489.
- MARASCUILLO, L. A. y MCSWEENEY, M. (1977): *Nonparametric and distribution-free methods for the social sciences* (Monterrey. California, Brooks Cole).
- PASTORE, N. (1950): Some comments on the use and misuse of the chi-square test, *Psychological Bulletin*: 47, pp. 338-340.
- PETERS, C. C. (1950): The misuse of chisquare. A reply to Lewis and Burke, *Psychological Bulletin*: 47, pp. 331-337.
- REYNOLDS, H. T. (1977): *The analysis of nominal data* (New York, Sage).
- SMITH, J. E. K. (1976): Analysis of qualitative data, *Annual Review Psychological*: 27, pp. 487-499.
- WALHS, J. E. (1962): *Handbook of nonparametric statistics* Volumen I (Princeton. New Jersey, Van Nostrand)...
 (1965): *Handbook of nonparametric statistics*, Volumen II (Princeton. New Jersey, Van Nostrand).
 (1968): *Handbook of nonparametric statistics*, Volumen III (Princeton. New Jersey, Van Nostrand).

SUMARIO: El presente artículo es la segunda parte del publicado con el mismo título general en el número anterior de esta misma revista. Allí tratamos las muestras del mismo tamaño y ahora nos centramos en el estudio de muestras de tamaño diferente.

Tanto la metodología utilizada como los conceptos básicos implicados están allí definidos.

El objetivo básico de esta segunda parte es analizar si se mantienen las conclusiones entonces obtenidas o si, por el contrario existen peculiaridades metodológicas e interpretativas en el análisis de datos.

Simulamos tres posibles situaciones muestrales: las dos primeras de carácter restringido y la tercera (variación aleatoria) de carácter más general.

Terminamos ofreciendo unas conclusiones estadístico-metodológicas para la investigación en ciencias humanas, referidas tanto al estadístico "ji-cuadrado" como a los diferentes valores de correlación propuestos.